BAYESIAN FORMULATIONS OF MULTIDIMENSIONAL BARCODE INVERSION*

MATTHEW M. DUNLOP[†], CHARLES M. ELLIOTT[‡], VIET HA HOANG[§], AND ANDREW M. STUART[†]

Abstract. A pair of Bayesian approaches to the reconstruction of binary functions in $\mathbb{R}^d$, $d = 2, 3$, is adopted; one is based on a Ginzburg-Landau penalized Gaussian prior, the other on a Bayesian level set formulation. For the Ginzburg-Landau approach a link is made to classical methods based on least squares constrained to piecewise constant functions, with a total variation regularization term which penalizes interfaces. This link rests on a careful choice, and scaling with respect to noise amplitude, of the prior and of the Ginzburg-Landau penalty. The key technical tool used to establish links between the Bayesian and classical approaches is the use of $\Gamma-$limits to study the MAP estimator. Furthermore, the parameter choices and scalings are shown, by means of numerical experiments, to lead to posterior concentration around a point which adequately encapsulates the truth. However, the Bayesian level set method is also shown to produce reconstructions of similar quality, at considerably lower computational cost, suggesting that the Bayesian level set method may be a viable alternative to total variation regularization for certain problems. The numerical studies are conducted using function-space MCMC.

Key words. Ginzburg-Landau, Gamma-convergence, level set method, Bayesian inverse problems, imaging, binary reconstruction.

AMS subject classifications. 35J35, 62G08, 62M40, 94A08.

1. Introduction.

The need to quantify uncertainty in predictions is of growing importance throughout the sciences and engineering. In this context the Bayesian approach to predictive science is very persuasive. On the other hand in most fields there is typically substantial development of classical, non-probabilistic, methodologies for making predictions, which predates Bayesian thinking. It is thus of interest to reconcile the Bayesian and classical perspectives, in order to facilitate transfer of ideas between the two communities. This idea was highlighted in the paper [10] where it was demonstrated that a number of classical algorithms in numerical analysis, such as quadrature rules, could be viewed as the conditional mean of a Bayesian inference problem. This perspective on the algorithms of numerical analysis has been developed quite broadly, including in the numerical solution of differential equations [23, 6, 21] and in numerical linear algebra [14]. In the context of inverse problems the link between Bayesian and classical approaches is well-understood in the setting of Gaussian random field priors: the Bayesian maximum a posteriori (MAP) estimator [17] is then the solution of a Tikhonov-Phillips regularized least squares problem [11] in which the regularization penalty is the square of the Cameron-Martin norm of the prior [8]; the Cameron-Martin space is necessarily a Hilbert space and the regularization is of Sobolev type. However when more complex priors are used the connection between classical and Bayesian perspectives is more subtle, even for linear inverse

*Submitted to the editors DATE.

Funding: The research of CME was partially supported by the Royal Society via a Wolfson Research Merit Award, the work of AMS by DARPA and the work of CME and AMS by the EPSRC programme grant EQUIP. MMD was partially supported by the EPSRC MASDOC Graduate Training Program. VHH gratefully acknowledges the MOE AcRF Tier 1 grant RG30/16.

[†]Computing & Mathematical Sciences, California Institute of Technology, USA, 91125 (mdunlop@caltech.edu, astuart@caltech.edu).

[‡]Mathematics Institute, University of Warwick, UK, CV4 7AL (c.m.elliott@warwick.ac.uk).

[§]Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371 (vhhoang@ntu.edu.sg).

problems, and there are interesting questions revolving around the interpretation of the MAP estimator and the conditional mean [3]. Many problems in image processsing require the reconstruction of piecewise continuous functions from noisy blurred data and classical methodologies for approaching this problem penalize the size of the interfaces between different smooth components of the solution. Helin and Lassas [13] recently showed how the Mumford-Shah functional [20] arises as the penalty term in a MAP estimator using hierarchical Gaussian priors for the linear inverse problem; their work builds on the framework suggested by Calvetti and Somersalo for the recovery of blocky images [4, 5]. However, obtaining total variation (TV) regularization as a straightforward MAP estimator, without hierarchical approaches, is problematic. The paper [18] demonstrates this phenomenon and shows how slightly different Besov space regularizations are more natural than TV regularization when one studies Bayesian MAP estimation of linear inverse problems.

In this paper we consider linear inverse problems for piecewise constant functions in two and three dimensions. Classical regularization methods applied to this problem are discussed in [2]. We introduce a non-Gaussian prior, found by applying a Ginzburg-Landau type penalty to a Gaussian random field prior. With an appropriate scaling of parameters in that prior, in terms of the small observational noise, we show that the resulting MAP estimator is closely related to minimization of the natural least squares functional, over the space of piecewise constant functions, penalized by the total variation of the function; more precisely we show that the $\Gamma-$limit of the MAP estimator is, in the small noise limit, given by this total variation regularized problem. Carrying this program out is of interest because it links a Bayesian methodology with a classical one, and sheds light on how to choose prior models in order to make this link concrete. We work entirely in two and three space dimensions and as a result need to use a high order generalization of the Modica-Mortola [19] regularization of TV in one dimension. The reason for this is the constraint that the Gaussian measure be supported on continuous functions meaning that the Cameron-Martin space is necesssarily smaller than H^1 in two or more dimensions; we use a generalization of Modica-Mortola to higher dimensions contained in [15]. Numerical results are presented for this non-Gaussian prior, and compared with an alternative approach, the Bayesian level set method [16]. The level set approach is shown to be superior from the point of view of sampling efficiently, and also produces similar quality of reconstruction; however the infimum of its MAP estimator functional is not attained [16] and, in addition, there is no direct link to TV regularization.

In section 2 we state the inverse problem of interest, and formulate it in terms of Bayesian inversion, explaining our two choices of non-Gaussian prior, and in both cases stating a well-posedness result for the posterior distribution. Section 3 characterizes the MAP estimator for the Ginzburg-Landau penalized Gaussian prior, and in particular demonstrates appropriate scaling of the prior required to obtain the desired $\Gamma-$limit in the small noise regime. The MAP estimator can be a computationally expedient way of studying Bayesian posterior distributions. The arguments in [16] show that the MAP estimator functional for the level set method does not exist, essentially because infimizing sequences tend to zero but the point $v = 0$ does not attain the infimum. However the MAP estimator for the Ginzburg-Landau approach has an interesting interpretation, which we explore in section 3, revealing connections to classical TV regularization of the inverse problem (1). In section 4 we describe numerical results which illustrate the foregoing discussion; related one dimensional numerical results may be found in [22]. We conclude in section 5, and Appendix A contains proofs of the main results.

2. Bayesian Formulations of the Inverse Problem. Let D be the unit cube $(0,1)^d \subset \mathbb{R}^d$. Let $X = C_{\#}(\bar{D}, \mathbb{R})$ where $\#$ denotes restriction to periodic functions. Let $K : L^1(D) \to \mathbb{R}^J$ be a bounded linear operator. As X is continuously embedded in $L^1(D)$, K is also a bounded continuous operator on X. We assume that our observational data is in the form of the image of the unknown function u under K with small additive noise:

$$y = Ku + \varepsilon^c \eta; \tag{1}$$

here ε and c denote constants, with $\varepsilon \ll 1$ and $c > 0$. We want to reconstruct u given the prior information that it takes values in the finite set $\{\pm 1\}$. The theory and numerical methods in this paper easily extend to globally Lipschitz nonlinear forward maps from $L^1(D)$ into $\mathbb{R}^J$.

2.1. High Order Ginzburg-Landau Formulation.

2.1.1. Prior. We construct a prior that is supported on continuous functions, but concentrates near to values ± 1. We achieve this by working with a measure absolutely continuous with respect to a Gaussian random field. Fix constants $\delta, \tau > 0$, $q \geq 0$ and $a_1, a_2, a_3 \in \mathbb{R}$. We define our Gaussian measure $\mu_0 = \mathsf{N}(0, C)$ on the Hilbert space $H := L^2(D)$ where C is the covariance operator defined implicitly by its inverse $C^{-1} : H^4_{\#}(D) \to H$ given by the identity $f = C^{-1}u \in H$ for $u \in H^4_{\#}(D)$ and

$$f = \delta\varepsilon^{-2a_1}\Delta^2 u - q\delta\varepsilon^{-2a_2}\Delta u + \tau^2\delta\varepsilon^{-2a_3}u.$$

It follows that E, the Cameron-Martin space of μ_0, is $H^2_{\#}(D)$ endowed with a norm

$$\|u\|_E^2 := \langle C^{-\frac{1}{2}}u, C^{-\frac{1}{2}}u\rangle = \delta \int_D \big(\varepsilon^{-2a_1}|\triangle u|^2 + q\varepsilon^{-2a_2}|\nabla u|^2 + \tau^2\varepsilon^{-2a_3}u^2\big)dx$$

where $\langle\cdot,\cdot\rangle$ denotes the standard $L^2(D)$ inner-product. Thus the prior is supported in X; see Lemma 6.25 of [24].

Now fix constants $r, b > 0$ and define the prior probability measure ν_0 on X via the Radon-Nikodym derivative

$$\frac{d\nu_0}{d\mu_0} = \frac{1}{Z_0}\exp\left(-\frac{r}{\varepsilon^b}\int_D \frac{1}{4}\big(1 - u(x)^2\big)^2 dx\right). \tag{2}$$

The normalization Z_0 is chosen so that ν_0 is a probability measure. Since the Gaussian measure μ_0 is supported on continuous functions in dimensions 2 and 3, so is the non-Gaussian measure ν_0. Furthermore, since $r, b > 0$ and $\varepsilon \ll 1$, this measure will concentrate on functions taking values close to ± 1.

2.1.2. Likelihood. Let η be a normal random variable in $\mathbb{R}^J$, $\eta \sim \mathsf{N}(0, \Sigma)$ with $\Sigma \in \mathbb{R}^{J\times J}$ is the positive-definite covariance of the noise. Then the random variable $y|u$, defined by (1), is distributed as the Gaussian $\mathsf{N}(Ku, \varepsilon^{2c}\Sigma)$ and this defines the likelihood.

2.1.3. Posterior. We let $\nu^y(du)$ denote the probability of the conditioned random variable $u|y$. Recall the Hellinger distance between measures μ and μ', defined with respect to any common reference measure μ_0 (but independent of it) and given by

$$d_{\mathrm{hell}}(\mu, \mu') = \sqrt{\left(\int_X \left(\sqrt{\frac{d\mu}{d\mu_0}} - \sqrt{\frac{d\mu'}{d\mu_0}}\right)^2 d\mu_0\right)}.$$

The following is a straightforward application of the theory in [9]:

PROPOSITION 2.1. *The posterior probability ν^y is a probability measure supported on X and determined by*

$$\frac{d\nu^y}{d\nu_0} = \frac{1}{Z}\exp\left(-\frac{1}{2\varepsilon^{2c}}\left|\Sigma^{-1/2}(y-Ku)\right|^2\right)$$

where $Z \in (0,\infty)$ is the normalization constant that makes ν^y a probability measure. Furthermore, the posterior measure ν^y is locally Lipschitz continuous with respect to $y \in \mathbb{R}^J$; more precisely: if $|y| < \rho$ and $|y'| < \rho$ for a constant $\rho > 0$ then there is a constant $C = C(\rho)$ such that

$$d_{\rm hell}(\nu^y, \nu^{y'}) \le C(\rho)|y-y'|.$$

2.2. Level Set Formulation.

2.2.1. Prior. We define the thresholding function $S : \mathbb{R} \mapsto \{-1, 0, +1\}$ by

$$S(v) = 1,\ v > 0, \quad S(0) = 0 \quad \text{and} \quad S(v) = -1,\ v < 0.$$

We assume that $u = S(v)$ and place the Gaussian prior μ_0 on v; this prior on v is supported on the function space X. It induces a prior on u by push forward under S and furthermore, under this prior, $u \in \{\pm 1\}$ D a.e., with probability 1; this is because the level sets of the Gaussian random field v have Lebesgue measure zero [16]. We work with v as our unknown for the purposes of inversion, noting that u is easily recovered by application of $S(\cdot)$. In particular draws from the induced prior on u may be created by writing

$$u(x) = \mathbb{1}_{v>0}(x) - \mathbb{1}_{v<0}(x)$$

with $v \sim \mu_0$.

2.2.2. Likelihood. Let η be a normal random variable in $\mathbb{R}^J$, $\eta \sim \mathsf{N}(0, \Sigma)$, Σ as before. Using the fact that $u = S(v)$ it follows that the random variable $y|v$ is distributed as the Gaussian $\mathsf{N}\big(KS(v), \varepsilon^{2c}\Sigma\big)$ and this defines the likelihood.

2.2.3. Posterior. Application of the theory in [16] gives the following:

PROPOSITION 2.2. *The posterior probability μ^y is a probability measure supported on X and determined by*

$$\frac{d\mu^y}{d\mu_0} = \frac{1}{Z}\exp\left(-\frac{1}{2\varepsilon^{2c}}\left|\Sigma^{-1/2}\big(y-KS(u)\big)\right|^2\right)$$

where $Z \in (0,\infty)$ is the normalization constant that makes μ^y a probability measure. Furthermore, the posterior measure μ^y is locally Lipschitz continuous with respect to $y \in \mathbb{R}^J$; more precisely: if $|y| < \rho$ and $|y'| < \rho$ for a constant $\rho > 0$ then there is a constant $C = C(\rho)$ such that

$$d_{\rm hell}(\mu^y, \mu^{y'}) \le C(\rho)|y-y'|.$$

3. MAP Estimators for Ginzburg-Landau Posterior. Recall the Cameron-Martin space E of the Gaussian measure μ_0 on X is $H^2_{\#}(D)$ with the norm given by given by

$$\|u\|_E^2 = \frac{\delta}{\varepsilon^{2a_1}}\|\triangle u\|^2_{L^2(D)} + \frac{\delta q}{\varepsilon^{2a_2}}\|\nabla u\|^2_{L^2(D)} + \frac{\delta\tau^2}{\varepsilon^{2a_3}}\|u\|^2_{L^2(D)}.$$

Now define $\Psi : X \to \mathbb{R}^+$ by

$$\Psi(u) = \frac{r}{\varepsilon^b} \int_D \frac{1}{4}\big(1 - u(x)^2\big)^2 dx \tag{3}$$

and $\Phi : X \times \mathbb{R}^J \to \mathbb{R}^+$ by

$$\Phi(u, y) = \Psi(u) + \frac{1}{2\varepsilon^{2c}} |\Sigma^{-1/2}(y - Ku)|^2. \tag{4}$$

We define the functional J^ε by

$$J^\varepsilon(u) = \begin{cases} \frac{1}{2}\|u\|_E^2 + \Phi(u; y) & \text{if } u \in E, \\ \infty & \text{if } u \notin E. \end{cases}$$

This is the Onsager-Machlup functional associated with measure ν^y. For $\rho > 0$, let $B^\rho(z)$ be the ball centred at $z \in X$ with radius ρ and define

$$z^\rho = \operatorname{argmax}_{z \in X} \nu^y(B^\rho(z)).$$

Following Dashti et al. [8], we define a MAP estimator as follows. Intuitively this definition captures the idea that the MAP estimator locates points in X at which arbitrarily small balls will have maximal probability.

DEFINITION 3.1. *A point $\bar{z} \in X$ is a MAP estimator for the posterior measure ν^y if*

$$\lim_{\rho \to 0} \frac{\nu^y(B^\rho(\bar{z}))}{\nu^y(B^\rho(z^\rho))} = 1.$$

Then we have the following result, which follows from Theorem 3.5 of Dashti et al. [8].

PROPOSITION 3.2. *There exists a MAP estimator for the posterior measure ν^y which is a minimizer of the functional J^ε.*

This demonstrates the role of the Onsager-Machlup functional. We now study its Γ limit as $\varepsilon \to 0$. The functional $J^\varepsilon(u)$ can be written as

$$J^\varepsilon(u) = \varepsilon^{-2a_1-3} I^\varepsilon(u), \tag{5}$$

where

$$\begin{aligned} I^\varepsilon(u) =& \frac{1}{2}\delta\varepsilon^3 \|\triangle u\|_{L^2(D)}^2 + \frac{1}{2}\delta q \varepsilon^{3+2(a_1-a_2)} \|\nabla u\|_{L^2(D)}^2 + \frac{1}{2}\delta\tau^2\varepsilon^{3+2(a_1-a_3)} \|u\|_{L^2(D)}^2 \\ &+ r\varepsilon^{3+2a_1-b} \int_D \frac{1}{4}\big(1 - u(x)^2\big)^2 dx + \frac{1}{2}\varepsilon^{3+2a_1-2c} \big|\Sigma^{-1/2}(y - Ku)\big|^2. \end{aligned}$$

As $\varepsilon > 0$ the critical points of J^ε and I^ε coincide, therefore in what follows we will consider I^ε. We consider the case where

$$a_2 - a_1 = 1, \quad 3 + 2a_1 - b = -1, \quad 3 + 2a_1 - 2c = 0, \quad 3 + 2(a_1 - a_3) = a > 0. \tag{6}$$

Then the functional $I^\varepsilon(u)$ becomes

$$\begin{aligned} I^\varepsilon(u) = \int_D \left(\frac{1}{2}\delta\varepsilon^3 |\triangle u|^2 + \frac{1}{2}\delta q\varepsilon |\nabla u|^2 + \frac{r}{4\varepsilon}\big(1 - u(x)^2\big)^2 + \delta\tau^2\varepsilon^a u(x)^2 \right) dx& \\ + \frac{1}{2}\big|\Sigma^{-1/2}(y - Ku)\big|^2& \end{aligned}$$

and $I^\varepsilon(u) = +\infty$ when $u \in H \setminus H^2_\#(D)$. Note that we may achieve the critical scalings (6) with many choices of the parameters. For example $a_2 = 0, a_1 = -1, b = 2$ and $c = \frac{1}{2}$. Note, however, that the constraint $c > 0$ (the small noise assumption) means that $2a_1 + 3 > 0$.

We study the Γ convergence of the functional I^ε, basing our analysis on the work of Hilhorst et al [15]. We define the functional

$$e^\delta(U) = \int_{-\infty}^{\infty} \left(\frac{1}{2}\delta(U''(t))^2 + \frac{q}{2}\delta(U'(t))^2 + \frac{r}{4}(1 - U(t)^2)^2\right) dt;$$

and the constant

$$P^\delta = \inf_{U \text{ odd}} e^\delta(U).$$

We then have the following theorem:

Theorem 3.3. *Define*

$$I_0^\delta = \frac{1}{2}\int_D P^\delta |\nabla u| + |\Sigma^{-1/2}(y - Ku)|^2 dx, \quad if\ u \in BVC(D),$$

where $BVC(D) = \{\psi \in BV(D) :\ \psi(D) \subset \{\pm 1\}\}$. *Then*

$$I_0^\delta = \Gamma - \lim_{\varepsilon \to 0} I^\varepsilon$$

in the strong $L^1(D)$ *topology.*

We present the proof in Appendix A. Since $2a_1 + 3 > 0$, (5) together with the preceding $\Gamma-$limit theorem suggest that, when $\varepsilon \ll 1$, the measure will approximately concentrate on a single point close to a minimizer of I_0^δ. Our numerical results will support this conjecture.

4. Numerical Results. Markov Chain Monte Carlo (MCMC) simulations may be used to sample the measures ν^y, μ^y defined above. These samples can then be used to produce point estimates for the unknown fields, by calculating, for example, their mean or the sign of their mean. We compare the cost of sampling versus the quality of reconstruction with these point estimates, for both the Ginzburg-Landau and level set formulations. Preliminary numerical results for the one dimensional analogue of the problem studied here may be found in [22]. In one dimension the natural regularization of TV is through the Modica-Mortola functional [19] and the resulting Cameron-Martin space is $H^1_\#$.

4.1. Sampling and Setup. We outline the preconditioned Crank-Nicolson (pCN) algorithm, which may be used to sample a measure σ of the form

$$\frac{d\sigma}{d\sigma_0}(u) = \frac{1}{Z}\exp\left(-A(u;y)\right), \quad \sigma_0 = N(0, C).$$

Taking $\sigma_0 = \mu_0$ and $A(u;y) = \Phi(u;y)$ provides the Ginzburg-Landau posterior $\sigma = \nu^y$. Similarly, taking $\sigma_0 = \mu_0$ and

$$A(u;y) = \frac{1}{2\varepsilon^{2c}}\left|\Sigma^{-1/2}\big(y - KS(u)\big)\right|^2$$

provides the level set posterior $\sigma = \mu^y$. The pCN algorithm was introduced in [1]; see [7] for a review. It has the advantage that, unlike the standard Random Walk

Algorithm 1 Preconditioned Crank-Nicolson

1. Fix $\beta \in (0,1]$. Choose an initial state $u^{(0)} \in X$ and set $n = 0$.
2. Propose a state
$$v^{(n)} = (1-\beta^2)^{\frac{1}{2}} u^{(n)} + \beta\xi^{(n)}, \quad \xi^{(n)} \sim N(0, C).$$
3. Set $u^{(n+1)} = v^{(n)}$ with probability
$$\alpha(u^{(n)}, v^{(n)}) = \min\{1, \exp(A(u^{(n)}; y) - A(v^{(n)}; y))\},$$
or else set $u^{(n+1)} = u^{(n)}$.
4. Set $n \mapsto n+1$ and go to 2.

Metropolis MCMC algorithm, its rate of convergence to equilibrium can be bounded below independently of mesh resolution [12].

For the Ginzburg-Landau formulation, we make a choice of parameters in the prior covariance operator C such that the relations (6) hold; the MAP estimator for ν^y then approximates the minimizer of I_0^δ as given in Theorem 3.3. Specifically, we make the choices $a_1 = 0$, $a_2 = 1$, $a_3 = 0$, $b = 4$, $c = 3/2$, $\varepsilon = 0.01$, $\delta = 0.01$, $q = 0.1$ and $\tau = 1$. Additionally, we set $r = 1$.

For the level set formulation we make the same choices of prior parameters as for the Ginzburg-Landau formulation, except we set $\delta = 1$, $q = 0$ and $\tau = 50$. Note that in general we need not insist on the parameters being related via (6) for the level set formulation; this is because, unlike the Ginzburg-Landau formulation, there is no MAP estimator whose properties we are seeking to control via parameter choices.

Two fields, referred to as Truth A and Truth B, are used to generate the data and are presented in Figure 1. They are observed on a uniform grid of 15×15 points, and these observations are corrupted by additive Gaussian noise with standard deviation $\varepsilon^c = 0.001$, as in equation (1). Truth A and Truth B are generated on a square mesh of 2^{16} points, and sampling is performed on a mesh of 2^{14} points to avoid an inverse crime. In all cases we generate 10^6 samples, and discard the first 5×10^5 samples as burn-in when computing means. We take $\beta = 0.002$ for Ginzburg-Landau simulations, and $\beta = 0.02$ for level set simulations; these choice are made in order to balance acceptance and mixing rates.

4.2. Results. In Figure 2 we present sample means associated with both Truth A and Truth B. General quality of reconstruction is similar for both the Ginzburg-Landau and level set priors, though only the latter produces binary fields. Additionally, as a consequence of approximating a piecewise constant function by Fourier modes – the eigenfunctions of the covariance operator – Gibbs effects are present in the Ginzburg-Landau reconstructions. For Truth A, the effect of the observation grid not aligning with all of the inclusions is reflected in both reconstructions, by for example the squaring of one half of the circle. Such effects are also seen in the reconstructions of Truth B, though the overfitting to the observation points appears to be slightly less prominent in the level set reconstruction.

In terms of computational cost, every chain requires an evaluation of $A(\cdot;\cdot)$ per sample. Due to the presence of an extra integral term, this evaluation will typically be more expensive for the Ginzburg-Landau model than the level set model; for the simulations performed here, evaluation of $A(\cdot;\cdot)$ was approximately twice as expensive

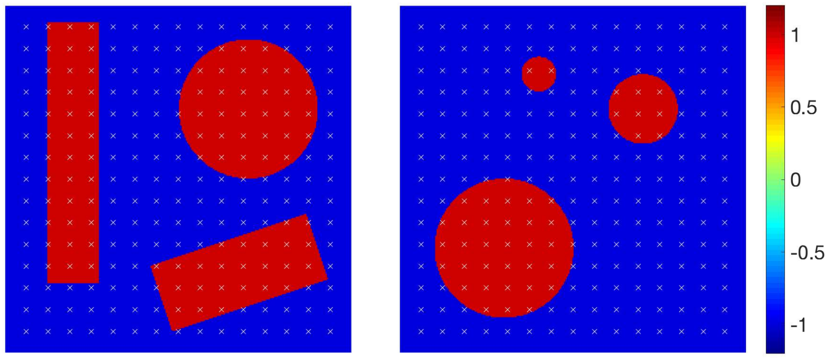

Fig. 1. *The two true fields used for inversion; the field on the left will be referred to as Truth A and the field on the right as Truth B. The* 15×15 *grid of observation points is shown in each figure.*

for the Ginzburg-Landau model than for the level set model.

A more significant increase in computational cost arises from the statistical properties of the chains. In Figure 3 we show the evolution of the local acceptance rates of proposed states. These figures suggest that the Ginzburg-Landau chains have not reached equilibrium until after at least 5×10^5 samples, whereas the level set chains converge much earlier. This is illustrated in Figure 4, which shows a selection of samples early in the chains for Truth B. After 10000 samples the three inclusions have already been identified by the level set chain, however after 50000 samples the Ginzburg-Landau chain has only started to identify a second inclusion. Thus, even though for both models we produced the same number of samples, it would have sufficed to terminate the level set chains much earlier, significantly reducing the computational cost.

Another observation to make from Figure 3 is that the acceptance rates for the Ginzburg-Landau chains are much lower than those for the level set chains, despite the jump parameter β being one tenth of the size. To understand why this is the case, note that the measure ν^y can informally be thought of as having Lebesgue density proportional to $\exp(-J^\varepsilon(u)) = \exp(-\varepsilon^{-2a_1-3} I_\delta^0(u))$. Thus for small ε the probability mass is concentrated in a small neighborhood of critical points of $I^\varepsilon \approx I_\delta^0$. The MCMC simulations for ν^y could hence be viewed as a form of derivative-free optimization for the functional J^ε.

5. Conclusions. We have studied Bayesian piecewise constant reconstruction for the linear inverse problem. Two approaches are considered, one based on Ginzburg-Landau penalization, the other on a level set formulation. We have determined a form of prior, and scalings of it with respect to assumed small observational noise, which ensure that the MAP estimation functional for the Ginzburg-Landau approach is close to a total variation penalized least squares problem. With this scaling on the prior the posterior measure concentrates on the truth. We have also demonstrated numerically that the Bayesian level set approach obtains a similar level of accuracy in reconstruction, at considerably less computational cost.

Appendix A. Proofs of Main Results.

Proof of Proposition 3.2. Throughout this proof C is a universal constant whose value

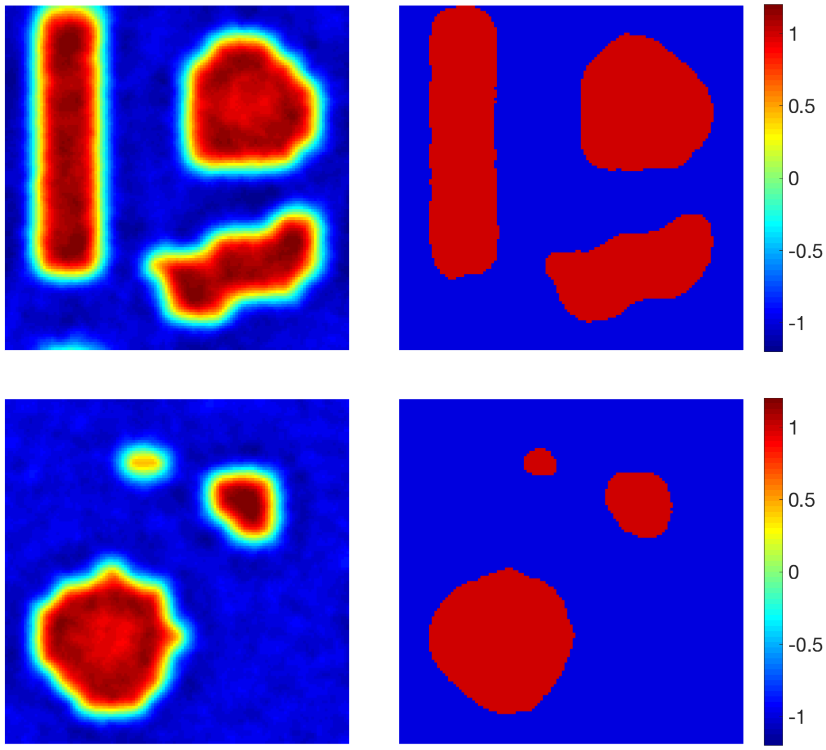

FIG. 2. *Sample means for Truth A (top) and Truth B (bottom). Monte Carlo approximations to $\mathbb{E}^{\nu^y}(u)$ (Ginzburg-Landau, left) and $S\big(\mathbb{E}^{\mu^y}(u)\big)$ (level set, right) are shown*

may change between occurences. To apply Theorem 4.12 from [8], we need to show that the function $\Phi(\cdot, y)$ is bounded from below, is locally bounded from above and is locally Lipschitz. We note that $\Phi(\cdot, y)$ is always non-negative so is bounded from below. If $\|u\|_X = \max_{x\in\bar{D}} |u(x)| \le \rho$ then we may bound $|\Phi(u, y)|$ by a constant depending on ρ i.e. $\Phi(\cdot, y)$ is locally bounded. For the local Lischiptzness, we have

$$\Phi(u,y) - \Phi(v,y) = \frac{r}{4\varepsilon^b}\int_D (2 - u(x)^2 - v(x)^2)(u(x)+v(x))(u(x)-v(x))dx + \\ \frac{1}{2\varepsilon^{2c}}\langle \Sigma^{-1/2}(2y - Ku - Kv), \Sigma^{-1/2}K(v-u)\rangle$$

Assume that $\|u\|_X \le \rho$ and $\|v\|_X \le \rho$. Then, since K is a bounded linear operator on $L^1(D)$,

$$\begin{aligned}|\Phi(u,y) - \Phi(v,y)| &\le C\int_D |u(x)-v(x)|dx + C|K(v-u)| \\ &\le C\|u-v\|_{L^1(D)} \le C|D|^{1/2}\|u-v\|_{L^2(D)} \le C\|u-v\|_X.\end{aligned}$$

The desired result follows. ☐

Proof of Theorem 3.3. We adapt the proof of Hilhorst et al. to allow for periodic boundary conditions and the additional L^2 norm appearing in the functional to be

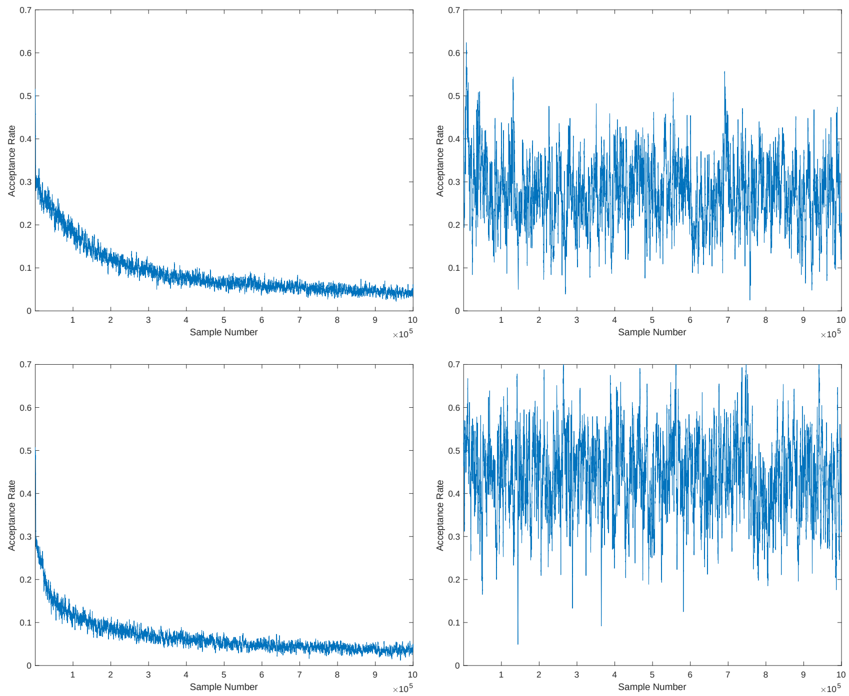

Fig. 3. *The evolution of the acceptance rate of proposals for the Ginzburg-Landau (left) and level set (right) MCMC chains, for Truth A (top) and Truth B (bottom). Acceptance rates are calculated over a moving window of* 1000 *samples.*

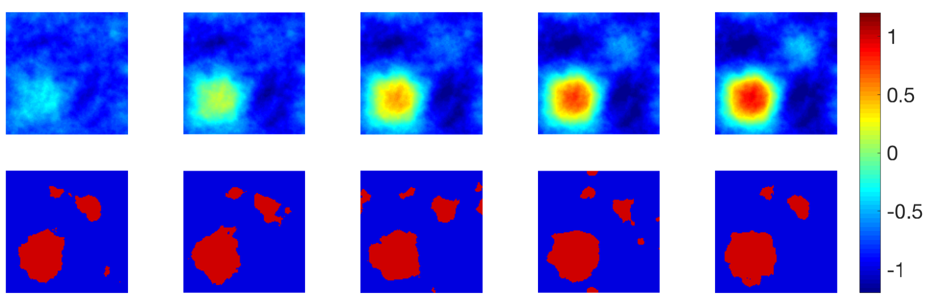

Fig. 4. *Examples of samples near the start of the chains for Truth B. Sample numbers 10000, 20000, 30000, 40000 and 50000 are shown from left-to-right for the Ginzburg-Landau chain (top) and the level set chain (bottom).*

infimized. From Hilhorst et al., we have that if $u^\varepsilon \to u$ in $L^1(D)$ then

$$\begin{aligned}\liminf_{\varepsilon\to 0} I^\varepsilon(u^\varepsilon) \geq \liminf_{\varepsilon\to 0} & \int_D \left(\frac{1}{2}\delta\varepsilon^3|\triangle u^\varepsilon|^2 + \frac{1}{2}\delta q\varepsilon|\nabla u^\varepsilon|^2 + \frac{r}{4\varepsilon}\big(1-u^\varepsilon(x)^2\big)^2\right)dx \\ & + \frac{1}{2}|\Sigma^{-1/2}(y-Ku^\varepsilon)|^2 \\ \geq{}& I_0^\delta(u).\end{aligned}$$

Now we show that for each $u \in L^1(D)$, there is a sequence $\{u^\varepsilon\} \subset H^2_\#(D)$ which converges strongly to u in $L^1(D)$ such that $\limsup_{\varepsilon\to 0} I^\varepsilon(u^\varepsilon) \le I^\delta_0(u)$. We first review the main points in the proof of Hilhorst et al. for functions $u \in H^2(D)$. Considering the case $I^\delta(u) < \infty$, without loss of generality, we assume that

$$u = \mathbb{1}_Q - \mathbb{1}_{\mathbb{R}^d\setminus Q}$$

where Q is a bounded domain, with $\partial Q \in C^\infty$ and $Q \subset\subset D$. The sign distance function is defined as

$$d(x) = \begin{cases} +\inf_{y\in\partial Q}|x-y| \text{ if } x \in Q \\ -\inf_{y\in\partial Q}|x-y| \text{ if } x \notin Q \end{cases}$$

Let N_h be an h neighbourhood of ∂Q (we choose h so that h is less than the distance between ∂Q and ∂D.) We choose a function $\eta \in C^2(\bar{D})$ such that $\eta(x) = d(x)$ for $x \in N_h$, $\eta(x) \ge h$ when $x \in Q\setminus N_h$ and $\eta(x) \le -h$ when $x \in D\setminus(Q\bigcup N_h)$. Let U be an odd minimizer of the functional $e^\delta(U)$ with $\lim_{t\to\infty} U(t) = 1$ and $\lim_{t\to-\infty} U(t) = -1$. We let

$$u^\varepsilon = U\left(\frac{\eta(x)}{\varepsilon}\right).$$

We note that $u^\varepsilon(x)$ is uniformly bounded pointwise and $u^\varepsilon(x) \to u(x)$ for all $x \in D$. From the Lebesgue dominated convergence theorem, $u^\varepsilon \to u$ in $L^1(D)$ and in $L^2(D)$. Thus

$$\lim_{\varepsilon\to 0} |\Sigma^{-1/2}(y - Ku^\varepsilon)|^2 = |\Sigma^{-1/2}(y - Ku)|^2.$$

and, since $a > 0$,

$$\lim_{\varepsilon\to 0}\int_D \delta\tau^2\varepsilon^a u^\varepsilon(x)^2 dx = 0.$$

To show that $\lim_{\varepsilon\to 0} I^\varepsilon(u) = I^\delta_0(u)$, we follow the approach of Hilhorst et al.. The integral

$$\int_D \left(\frac{1}{2}\delta\varepsilon^3|\triangle u^\varepsilon|^2 + \frac{1}{2}\delta q\varepsilon|\nabla u^\varepsilon|^2 + \frac{r}{4\varepsilon}\big(1 - u^\varepsilon(x)^2\big)^2\right)dx$$

is written as

$$\int_{D\setminus N_h} \left(\frac{1}{2}\delta\varepsilon^3|\triangle u^\varepsilon|^2 + \frac{1}{2}\delta q\varepsilon|\nabla u^\varepsilon|^2 + \frac{r}{4\varepsilon}\big(1 - u^\varepsilon(x)^2\big)^2\right)dx$$
$$+\int_{N_h} \left(\frac{1}{2}\delta\varepsilon^3|\triangle u^\varepsilon|^2 + \frac{1}{2}\delta q\varepsilon|\nabla u^\varepsilon|^2 + \frac{r}{4\varepsilon}\big(1 - u^\varepsilon(x)^2\big)^2\right)dx.$$

Using the exponential decay of U, U' and U'' at ∞ and $-\infty$, we deduce that the integral over $D \setminus N_h$ goes to 0 when $\varepsilon \to 0$ (note that $|\eta(x)|/\varepsilon > h/\varepsilon$ which goes to ∞ when $\varepsilon \to 0$ for $x \in D \setminus N_h$). The integral over N_h is shown to converge to $I^\delta(u)$ when $\varepsilon \to 0$.

To adapt this proof of Hilhorst et al. to functions with periodic boundary condition on D, we only need to choose the function η so that η is periodic and $\eta(x) \ge h$ when $x \in Q \setminus N_h$ and $\eta(x) \le -h$ when $x \in D \setminus (Q\bigcup N_h)$. Such a function can be constructed as follows. Let $\psi(x) \in C^\infty_0(D)$ be such that $\psi(x) = 1$ when x is in a neighbourhood of $Q\bigcup N_h$, and $0 \le \psi(x) \le 1$ for all $x \in D$. Let $\eta_1(x)$ be a smooth periodic function with $\eta_1(x) \le -h$ for all $x \in D$. Using the function η of Hilhorst et al., we define a new function

$$\bar{\eta}(x) = \psi(x)\eta(x) + (1 - \psi(x))\eta_1(x).$$

The function $\bar{\eta}(x)$ satisfies the requirement. □

REFERENCES

[1] A. Beskos, G. O. Roberts, A. M. Stuart, and J. Voss, *MCMC methods for diffusion bridges*, Stochastics and Dynamics, 8 (2008), pp. 319–350.

[2] C. Brett, C. M. Elliott, and A. S. Dedner, *Phase field methods for binary recovery*, in Optimization with PDE Constraints, Springer, 2014, pp. 25–63.

[3] M. Burger and F. Lucka, *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*, Inverse Problems, 30 (2014), p. 114004.

[4] D. Calvetti and E. Somersalo, *A Gaussian hypermodel to recover blocky objects*, Inverse Problems, 23 (2007), p. 733.

[5] D. Calvetti and E. Somersalo, *Hypermodels in the Bayesian imaging framework*, Inverse Problems, 24 (2008), p. 034013.

[6] O. A. Chkrebtii, D. A. Campbell, B. Calderhead, and M. A. Girolami, *Bayesian solution uncertainty quantification for differential equations*, Bayesian Anal., 11 (2016), pp. 1239–1267, https://doi.org/10.1214/16-BA1017, http://dx.doi.org/10.1214/16-BA1017.

[7] S. L. Cotter, G. O. Roberts, A. M. Stuart, D. White, et al., *MCMC methods for functions: modifying old algorithms to make them faster*, Statistical Science, 28 (2013), pp. 424–446.

[8] M. Dashti, K. J. Law, A. M. Stuart, and J. Voss, *MAP estimators and their consistency in Bayesian nonparametric inverse problems*, Inverse Problems, 29 (2013), p. 095017.

[9] M. Dashti and A. M. Stuart, *The Bayesian approach to inverse prob-lems*, arXiv preprint arXiv:1302.6989, (2014).

[10] P. Diaconis, *Bayesian numerical analysis*, Statistical Decision Theory and Related Topics IV, 1 (1988), pp. 163–175.

[11] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, vol. 375, Springer Science & Business Media, 1996.

[12] M. Hairer, A. M. Stuart, and S. J. Vollmer, *Spectral gaps for Metropolis-Hastings algorithms in infinite dimensions*, The Annals of Applied Probability, 24 (2014), pp. 2455–290, https://doi.org/10.1214/13-AAP982.

[13] T. Helin and M. Lassas, *Hierarchical models in statistical inverse problems and the Mumford–Shah functional*, Inverse problems, 27 (2011), p. 015008.

[14] P. Hennig, *Probabilistic interpretation of linear solvers*, SIAM Journal on Optimization, 25 (2015), pp. 234–260.

[15] D. Hilhorst, L. A. Peletier, and R. Schätzle, *γ-limit for the extended Fisher–Kolmogorov equation*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics, 132 (2002), pp. 141–162.

[16] M. A. Iglesias, Y. Lu, and A. M. Stuart, *A Bayesian level set method for geometric inverse problems*, Interfaces and Free Boundary Problems, (2016).

[17] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, vol. 160, Springer Science & Business Media, 2006.

[18] M. Lassas, E. Saksman, and S. Siltanen, *Discretization-invariant Bayesian inversion and Besov space priors*, Inverse Problems and Imaging, 3 (2009), pp. 87–122.

[19] L. Modica and S. Mortola, *Un esempio di γ-convergenza*, Boll. Un. Mat. Ital. B (5), 14 (1977), pp. 285–299.

[20] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Communications on pure and applied mathematics, 42 (1989), pp. 577–685.

[21] M. Schober, D. K. Duvenaud, and P. Hennig, *Probabilistic ODE solvers with Runge-Kutta means*, in Advances in Neural Information Processing Systems, 2014, pp. 739–747.

[22] I. Sivak, *Bayesian reconstruction of piecewise constant signals*, MSc. Dissertation, Warwick University, (2014).

[23] J. Skilling, *Bayesian solution of ordinary differential equations*, in Maximum Entropy and Bayesian Methods, Springer, 1992, pp. 23–37.

[24] A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numerica, 19 (2010), pp. 451–559.